Classification and Detection of Wind Turbine Pitch Faults Through SCADA Data Analysis

Jamie L. Godwin[1] and Peter Matthews[2]

[1,2]*School of Engineering and Computing Sciences, Durham University, Durham, DH1 3LE, United Kingdom*

j.l.godwin@durham.ac.uk
p.c.matthews@durham.ac.uk

ABSTRACT

The development of electrical control system faults can lead to increased mechanical component degradation, severe reduction of asset performance, and a direct increase in annual maintenance costs. This paper presents a highly accurate data driven classification system for the diagnosis of electrical control system faults, in particular, wind turbine pitch faults. Early diagnosis of these faults can enable operators to move from traditional corrective or time based maintenance policy towards a predictive maintenance strategy, whilst simultaneously mitigating risks and requiring no further capital expenditure. Our approach provides transparent, human-readable rules for maintenance operators which have been validated by an independent domain expert. Data from 8 wind turbines was collected every 10 minutes over a period of 28 months with 10 attributes utilised to diagnose pitch faults. Three fault classes are identified, each represented by 6000 instances in each of the testing and training sets. Of the turbines, 4 are used to train the system with a further 4 for validation. Repeated random sub-sampling of the majority fault class was used to reduce computational overheads whilst retaining information content and balancing the training and validation sets. A classification accuracy of 85.50% was achieved with 14 human readable rules generated via the RIPPER inductive rule learner. Of these rules, 11 were described as “useful and intuitive” by an independent domain-expert. An expert system was developed utilising the model along with domain knowledge, resulting in a pitch fault diagnostic accuracy of 87.05% along with a 42.12% reduction in pitch fault alarms.

1. INTRODUCTION

Maintenance costs for wind energy represent between 20-25% of total asset cost, of which, up to 75% is due to unscheduled maintenance (WWEA, 2012). This deters future investment, increases the cost of wind energy and as such, reduces the long term economic viability of wind energy. As corrective maintenance can be up to 40 times more expensive than a proactive strategy (Hatch, 2004) there is the potential for significant cost savings on wind turbine operations and maintenance (O&M) costs. For this reason, maintenance is moving from a “fail and fix” reactive approach to maintenance, to a “predict and prevent” strategy for maintenance (Levrat, et al 2008). Maintenance savings of 20-25% can be achieved using condition based maintenance (CBM) (Djurdjanovic, et al 2003), this is echoed by Wu & Clements-Croome (2005) who have shown the potential for preventive maintenance actions to be performed at between 10 times and 40 times that of corrective maintenance actions. However, uptake across all domains of prognostic technologies for the prediction of future failure modes has been slower than anticipated. It is believed that within the UK, CBM and prognostic technologies have only reached 10-20% penetration into industry (Moore & Starr, 2006). This is believed to be due to many factors, such as: the lack of transparency of some expert systems, the capital outlay required for data collection and analysis, the uncertainty and inaccuracy present within some techniques, staff training costs and no proven track record in similar domains. Whilst strategies such as reliability centred maintenance (RCM) can help optimise available maintenance resources, they are static nature in that they do not take into account the current level of asset degradation or external conditions. This means that whilst cost savings can be made through RCM (Niu, et al 2010), severe degradation is likely to go unnoticed for extended periods, causing secondary damage to auxiliary systems, reducing component efficiency and as a result, reduce overall return on investment for stakeholders. Due to as few as 20% of assets failing within the manufacturer prescribed times (Eti, et al. 2006), there is a need to move away from a static analysis towards a more dynamic, real-time approach to maintenance. Currently, maintenance is often seen by senior management as a cost minimisation exercise, rather than an attempt to maximise benefit (Marais & Saleh, 2009). This is due to the ease of quantifying the cost of maintenance, but not the benefit provided. This attitude towards maintenance means that most efforts to reduce annual maintenance expenditure result in a direct loss of availability or reduction in the quality of service provided (Gomez-Fernandez & Crespo-Marquez, 2009).

Typically, condition based monitoring is performed using high frequency data – acoustic emissions and vibration data – collected for the remote diagnosis and prognosis of the gearbox, generator and main bearing (Crabtree, 2010). However, being able to establish and track the development of a fault over longer lengths of time through utilising low frequency data is interesting as it provides feedback into the maintenance planning and scheduling process, enabling the optimisation of available resources, thereby reducing annual maintenance costs.

In this paper we present a new methodology for the development of a transparent expert system for the detection of wind turbine pitch faults utilising a data-intensive machine learning approach. This approach describes a classifier to determine the current condition of the pitch system on a wind turbine through analysis of low frequency SCADA data, and if a fault is observed within the pitch system, an expert system recommends the correct action to take depending upon its severity. Severe pitch faults requiring potential maintenance actions can then be presented to the maintenance operator whilst filtering out unnecessary information and reducing the cognitive load which is placed upon them. As the data utilised for this methodology is from a pre-existing SCADA system, no further sensors are required and no additional capital expenditure are incurred. This mitigates many of the risks associated with moving to a proactive maintenance strategy.

2. WIND TURBINE PITCH FAULTS

Wind turbine pitch faults a deviation of the blade pitch angle from a predefined optimum for a given wind speed are the most common fault mode to occur. As can be seen in Table 1, pitch faults account for over one third of all faults which are present within the SCADA system which are then presented to the maintenance operator. It is not uncommon for over 2,000 SCADA pitch fault alarms to occur over a year. However, less than 5% of these directly correlate to a maintenance action within the maintenance log; wasting available maintenance resources with undue inspection and analysis. As such, there is a need to develop a data-driven expert system to allow the encapsulation of the behaviours both during and immediately preceding a pitch fault so that maintenance operators can further understand the extent of the fault, the causation of the fault and the maintenance action required. Accurate identification of pitch faults is of particular interest to maintenance operators and decision makers, as these faults are often the result of the electrical control system, and not due to severe physical degradation of the pitch motors controlling the wind turbine blades. As such, when a pitch fault is identified, the potential exists to remotely reset the turbine pitch system. This enables the turbine to return to normal operating conditions, without the need for excessive downtime for the required inspection. As such the capacity for energy generation can be increased, with the potential risk of increased degradation on auxiliary components reduced. Should a mechanical fault be observed, this will then be diagnosed by the system presented in Section 4, enabling the effective scheduling and planning of maintenance activities.

3. RELATED WORK

Over recent years, interest in improving the efficiency of all aspects of the wind turbine life cycle has become of paramount importance to ensure the transition to a low carbon economy and ending the reliance on fossil fuels. As up to 25% of total cost is manifested as maintenance for a wind turbine, effective maintenance through condition based maintenance and proactive maintenance is essential to increasing global investment in wind energy, providing returns to stakeholders, reducing energy prices to consumers and ensuring continued reliable operation as transitions are made to the smart grid (Massoud Amin & Wollenberg, 2005). Prognosis of the wind turbine enables 5 key benefits to be provided to the operator as stated in Hameed et al (2009). They are:

1. The avoidance of premature failures - reducing secondary damage to components and catastrophic failures.
2. A reduction in maintenance costs - by reducing catastrophic failures and optimising inspection intervals.
3. The capability of remote diagnosis – essential due to the remote nature of offshore turbines.
4. An increase in generation capacity - prognosis enables maintenance to be performed at low wind speed to ensure maximal utilisation.
5. Optimised future designs - large quantities of data can be analysed to ensure new generation turbines are more reliable.

Typically, condition monitoring on a wind turbine focuses on the high value components; the gearbox, generator and main bearing (Crabtree, 2010). Strong prognostic capability is prevalent within the literature. For example, the work done by Lin & Zuo (2003) and Rafiee et al (2010) use wavelet filters to provide condition based maintenance on these components. Also, Wang & Makis (2009) utilise statistical methods (such as autoregressive models) to achieve similar aims. However, these techniques require the installation of various additional sensors to each turbine to be monitored, which can be costly to the operator. For a full review of high frequency techniques, please see the work of Jardine et al (2006) and Hameed et al (2009). Techniques utilising low frequency data, such as SCADA data, do exist.

Sub-system	Turbine 1	Turbine 2
Pitch	4035	4130
Weather	2775	2866
Inverter	1438	1751
Gearbox	504	374
Yaw	316	385
Communications	285	827
Total	9353	10333

Table 1 – SCADA alarms aggregated by subsystem over a 28 month period for 2 typical turbines.

Work done by Kim et al (2011) has shown the electrical system of a wind turbine is the most prone to establishing a fault condition. It has been shown that low frequency

SCADA data can be used in conjunction with both PCA and self-organising feature maps for fault classification. However, diagnosis to determine the turbine sub-assembly at fault is not performed. As such, whilst maintenance managers may know a turbine requires inspection, further manual analysis will be required to determine the cause of the fault. Chen (2011) utilises an artificial neural network for the automatic analysis of SCADA alarm data. This is utilised as a filter to determine which SCADA alarms are novel and warrant further analysis. Work done by Kusiak & Li (2011) has shown that a variety of data mining approaches (neural networks, ensembles of neural networks, the boosting tree algorithm, support vector machines and classification and regression trees) can be used to diagnose and prognose irregular wind turbine states. However, even when utilising many different data driven approaches, a low prognostic horizon (less than an hour) is achieved, and accuracy of the classification of fault instances ranges from 40% to 71% - a relatively weak classification.

4. METHODOLOGY

SCADA data from 8 wind turbines was collected over a period of 28 months and sampled every 10 minutes, across 190 channels. All of these wind turbines had pitch faults noted in their histories as assessed by their maintenance log book. There had been 243 recorded pitch faults across the 28 months for the 8 turbines, ranging from 6 – 60 pitch faults per turbine ($M = 30.38$, $SD = 16.16$). In total, 999,944 records were retrieved. This data was combined with SCADA alarm system data and maintenance log data to give a holistic overview of the condition of the turbine and so that pitch fault events could be analysed. Due to the inherent nature of the data acquisition, erroneous and missing values are common; these are manifested as implausible values, missing data and duplicate data. This is ascribed to malfunction of the sensors, mechanical systems, data collection systems and also imperfections within the SCADA system itself (Sainz, et al 2009). Due to these problems, the data must be cleansed before processing can take place. Both missing and duplicate values were removed; missing values cannot accurately describe the current state of the wind turbine, and duplicate values provide no additional information whilst simultaneously increasing computational overhead. Once this is complete, attribute selection is performed. Based upon the work of Chen et al (2011) and Kusiak & Verma (2011), 8 attributes were selected for their consistently strong performance for wind turbine pitch fault diagnosis. Chen et al (2011) presents an artificial neural network (ANN) approach to pitch fault diagnosis, however, the diagnosis accuracy (M = 42.07%; SD = 17.49%) is relatively poor and black box nature of the approach is difficult to interpret by domain experts and maintenance operators. Whilst the work of Kusiak & Verma (2011) provides improved accuracy for the prediction of wind turbine pitch faults (M = 76.70%; SD = 5.62%), the genetic algorithm used provides human readable rules which are not necessarily transparent or easy to interpret by operators. As such, the attributes chosen for the model based upon the work in the literature (Chen et al., 2011 and Kusiak & Verma, 2011) were:

- Average wind speed
- Maximum wind speed
- Blade 1 pitch motor torque maximum
- Blade 2 pitch motor torque maximum
- Average pitch motor torque
- Blade 1 pitch angle average
- Blade 2 angle average
- SCADA pitch fault alarm status

In conjunction with these attributes, 2 additional derived parameters were utilised based upon the work of Chen et al (2011). These are:

- The absolute difference in torque across pitch motors
- The absolute difference in blade angle position

These attributes were chosen as they fully encapsulate the current operating characteristics of the wind turbine pitch fault system. The feathering control strategy for variable pitch wind turbines is described in detail by Bianchi et al. (2006).

For a given wind speed, each blade should be set to a pre-determined pitch based upon the strategy employed by the individual turbine. The pitch of all the wind turbine blades should be identical, and as such, deviations in either pitch or torque across the blades can be used to identify the presence of a pitch fault. The wind speed and SCADA alarms status provide additional context to the classifier to aid in the classification accuracy.

Following this, the data was classified into three distinct groups; "No pitch fault", "Pitch fault developing" and "Pitch fault established". These represent the development of a fault over time within the wind turbine. By classifying the data in this way we can identify both the wind turbines which urgently require maintenance and also the turbines with a reduced remaining useful life (RUL). Maintenance logs were used to determine when pitch faults had been severe enough to warrant a maintenance action.

The SCADA data from the 48 hours preceding this maintenance action was used to describe the "Pitch fault established" class. The SCADA data prior to this where the SCADA-alarm for the pitch fault was active was used to describe the "Pitch fault developing" class. Finally, all other data was used to describe the "no pitch fault" class. Annual maintenance costs can then be reduced utilising this classification; either by scheduling further turbines into existing maintenance actions, or by pre-emptively scheduling those which require maintenance before they become inaccessible to external factors. Repeated random sampling with 20 samples was utilised to remove the majority class bias inherent within the data. As "No pitch fault" was the dominant class and the turbine remains in this state for a prolonged period, a data-driven classifier would be stronger if it encapsulates this class well and ignores the pitch faults. However, as the aim of the system is the quality of the rules which describe the behaviour of the pitch faults, it is essential that this bias is removed so that the minority fault classes are encapsulated and characterised effectively. Within our data, the imbalance was typically between 125 to 380 instances per fault instance. Whilst other minority oversampling techniques could have been used such as SMOTE, MSMOTE and FSMOTE (Garcia, et al 2012) no significant increase in rule accuracy was attained over using traditional repeated random sampling within our dataset. As such, the majority class was under sampled, and the minority class oversampled until the data was balanced. After the data had been pre-processed, the RIPPER propositional rule learning algorithm (Cohen & Singer, 1999) was used to generate order independent, distinct encapsulations of explicit knowledge from the dataset. This technique was chosen due to its transparent, human-readable nature; ensuring trust was placed in the derived rules. An example of rules generated by the RIPPER algorithm can be seen in the appendix. Although other techniques such as artificial neural networks can achieve high quality classifications, their "black box" nature makes them difficult to extract meaningful rules from. Similarly, although techniques such as clustering and instance-based classification seem intuitive, the high-dimensionality of the dataset and high levels of noise present means that decision regions are non-convex in nature and neither a high level of accuracy nor good quality of rules can be extracted from the system. Decision tree algorithms could have been utilised, however, each rule generated cannot be understood independently from the system, and as such, can be difficult to extract and encapsulate as a single unit of knowledge.

4.1. Ripper Algorithm

The RIPPER algorithm (Cohen, 1995), is an extension to the IREP algorithm proposed by Fürnkranz and Widmer (1994), utilizing reduced error pruning (REP) used in decision tree algorithms. However, where the rule induction from decision trees is done in a breadth-first manner (as per C4.5), rule induction is performed in a depth-first manner.

There are two main stages within the ripper algorithm as described by Cohen (1995). Firstly, the data is split into "growing" and "pruning" dataset, with two thirds typically used for growing. This is done by random partitioning of the data. After this, rules are grown. This is done by adding conditions to a rule (greedily) until it is 100% accurate (that is, it covers no negative instance in the growing dataset). This is done by maximizing Foil's information gain criterion (Quinlan, 1990):

$$FOIL(L,R) = t \cdot (\log_2 \frac{p_1}{p_1+n_1} - \log_2 \frac{p_0}{p_0+n_0}) \qquad (1)$$

Where L is the condition to be added to R, t is the number of positive instances covered by $R+L$, p_1 and p_0 are the number of positive instances covered by R and $R+L$ (respectively), and n_1 and n_0 are the number of negative instances covered by R and $R+L$ (respectively). This favours rules which have high accuracy and cover many positive instances.

Once the rule has been grown, it is pruned immediately. This is done within RIPPER by considering the removal of the final sequence of conditions from the rule that maximise rule value:

$$v^*(Rule, PrunePos, PruneNeg) = \frac{p-n}{P+n} \qquad (2)$$

Where p is the number of examples in $PrunePos$ covered by $Rule$ and n is the number of examples in $PruneNeg$ covered by $Rule$. This is done until no deletion increases the value of $v*$ (Cohen, 1999).

Once the rules have been generated, optimisation is performed. In this stage, for each rule which has been grown and pruned, two variants are produced; the replacement and the revision. The replacement is generated by growing and pruning a rule where the pruning stage is guided to maximize the accuracy of the entire rule base. The revision is generated by greedily adding conditions to the rule. The rule with the minimum descriptive length (of the original, revision or replacement rule) is then chosen for the final rule base. For completeness, the full pseudo-code for the RIPPER algorithm (Cohen, 1995) is presented in Figure 1 (Alpaydin, 2004).

```
Ripper(Pos,Neg,k)
    RuleSet ← LearnRuleSet(Pos,Neg)
    For k times
        RuleSet ← OptimizeRuleSet(RuleSet,Pos,Neg)
LearnRuleSet(Pos,Neg)
    RuleSet ← ∅
    DL ← DescLen(RuleSet,Pos,Neg)
    Repeat
        Rule ← LearnRule(Pos,Neg)
        Add Rule to RuleSet
        DL' ← DescLen(RuleSet,Pos,Neg)
        If DL'>DL+64
          PruneRuleSet(RuleSet,Pos,Neg)
          Return RuleSet
        If DL'<DL DL ← DL'
          Delete instances covered by Rule from Pos and Neg
    Until Pos = ∅
    Return RuleSet
PruneRuleSet(RuleSet,Pos,Neg)
    For each Rule ∈ RuleSet in reverse order
        DL ← DescLen(RuleSet,Pos,Neg)
        DL' ← DescLen(RuleSet-Rule,Pos,Neg)
        IF DL'<DL Delete Rule from RuleSet
    Return RuleSet
OptimizeRuleSet(RuleSet,Pos,Neg)
    For each Rule ∈ RuleSet
        DL0 ← DescLen(RuleSet,Pos,Neg)
        DL1 ← DescLen(RuleSet-Rule+
                ReplaceRule(RuleSet,Pos,Neg),Pos,Neg)
        DL2 ← DescLen(RuleSet-Rule+
                ReviseRule(RuleSet,Rule,Pos,Neg),Pos,Neg)
        If DL1=min(DL0,DL1,DL2)
                Delete Rule from RuleSet and
                        add ReplaceRule(RuleSet,Pos,Neg)
        Else If DL2=min(DL0,DL1,DL2)
                Delete Rule from RuleSet and
                        add ReviseRule(RuleSet,Rule,Pos,Neg)
    Return RuleSet
```

Figure 1 – Pseudo-code of the RIPPER algorithm (Alpaydin, 2004).

4.2. Training and Validation Turbine Selection

Of the 8 wind turbines, 4 were used for training with the remaining turbines used for validation. In order to ensure the robustness of the methodology against training turbine selection, all combinations of turbines for both training and validation were considered. In total, 70 combinations of varying training and validation turbines were created. These models created a Pareto surface compromising the trade-off between the number of rules and rule accuracy which were then presented to an independent domain expert, maintenance operator and decision maker. This allows for both a quantitative and qualitative analysis of these rules so that the causation and diagnosis of pitch faults could more effectively be understood. This enables operators to understand the underlying physical properties of pitch faults so that they can be trained or assisted to identify pitch faults before further damage occurs, which may lead to the turbine being shut down for corrective maintenance which is often expensive.

5. RESULTS

The RIPPER propositional rule learner was trained on 70 models so that the robustness of the methodology could be ensured. Pruning of the rule set was enabled to reduce the quantity of rules to prevent potential cognitive overload, and was utilized in conjunction with four optimization iterations with three fold partitioning of the data.

5.1. Robustness to data scarcity

As can be seen in Table 2, the quantity of data available for training influences the accuracy of the system developed and also the size of the rule base. In addition to the analysis described in section 4 (on the full dataset), analysis was also carried out on 4, 8, 12, 16, 20 and 24 months of available data to determine the influence of the quantity of data on both classification accuracy and size of the rule base.

Each analysis in Table 2 was performed on the full set of 70 models generated by choosing each combination of the 8 training and testing turbines. As such, in total 490 models were developed and assessed to analyse the robustness of the system to the quantity of training data which was available.

With regards to model accuracy, a Pearson product-moment correlation was used to assess the relationship between mean classification accuracy attained and the quantity of data used. Preliminary analyses showed this relationship to be linear with both variables normally distributed, as assessed by Shaprio-Wilk test ($p > .05$), and there were no outliers. There was a strong positive association between classification accuracy and the quantity of data, $r(7) = .91$; $p < .01$. This is also the case for maximum classification accuracy; $r(7) = .91$, $p < .01$, and minimum accuracy attained, $r(7) = .92$, $p < .01$.

This shows that there is a strong positive correlation between the quantity of data available for training and the accuracy of the RIPPER algorithm. As such, it was determined that as much data as is available should be utilised when performing rule extraction. It should be noted that the lower bound of classification accuracy increased by 17.62% from utilising 4 months of data to using the entire data set (28 months), whereas the upper bound increased by 1.68% over the same period. The mean accuracy increase was 3.48% over this period; however, the standard deviation of accuracies was reduced by 1.81% in this period. This indicates less sensitivity to the selection of wind turbines used for testing as more data to be available. This was to be expected.

With regards to the size of the rule base, another Pearson product-moment correlation was used to assess the

	4 Months	8 Months	12 Months	16 Months	20 Months	24 Months	Full Dataset
Mean Accuracy	77.29%	77.92%	78.53%	78.37%	78.89%	78.91%	80.77%
Max Accuracy	85.73%	86.18%	86.94%	86.53%	87.25%	87.73%	87.41%
Min Accuracy	51.41%	59.11%	56.49%	65.74%	63.75%	66.39%	69.03%
Accuracy SD	6.49%	5.72%	5.86%	5.15%	5.12%	5.32%	4.68%
Mean Rule Base	7.57 rules	9.10 rules	10.94 rules	12.87 rules	13.57 rules	14.77 rules	16 rules
Max Rule Base	15 rules	16 rules	23 rules	24 rules	32 rules	34 rules	38 rules
Min Rule Base	3 rules	4 rules	4 rules	4 rules	5 rules	6 rules	6 rules
Rule Base SD	2.42 rules	2.90 rules	4.10 rules	4.27 rules	4.76 rules	4.78 rules	5.77 rules

Table 2 – Robustness to data scarcity with descriptive statistics for classification accuracy and rule base size.

relationship between mean rule base size and the quantity of data used. Preliminary analyses showed this relationship to be linear with both variables normally distributed, as assessed by Shaprio-Wilk test ($p > .05$), and there were no outliers. There was a very strong positive association between the size of the rule and the quantity of data, $r(7) = .99$; $p < .01$. This was also the case for the maximum size of the rule base, $r(7) = .97$, $p < .01$, and also the case for the minimum size of the rule base, $r(7) = .97$, $p < .01$.

This correlation is to be expected based upon the behaviour of the RIPPER algorithm. However, due to this, a trade off does exist. Increasing the quantity of data available to the propositional rule learner would increase the quality of the classifier produced, but would also increase the quantity of rules generated for analysis. This is detailed below.

5.2. Model selection

Due to the higher mean accuracy and larger rule base variance attained by models utilising the full 28 months of data, this was chosen for further analysis. The accuracy of the classification for the full data models was in the range of 69.03% - 87.41% (*M* = 80.77%; *SD* = 4.68%), with the number of rules generated by each model being in the range of 6 – 38 (*M* = 16; *SD* = 5.77). After removal of the models which were dominated by those with stronger classification accuracy but the same number of rules, 21 models were eligible to be presented to independent domain experts and maintenance operators for critical analysis of the rules generated.

The 21 models developed had classification accuracy in the range of 69.99% - 87.41% (*M* = 82.70%; *SD* = 4.26%). Similarly, the quantity of rules generated were in the range of 6 – 38 (*M* = 16.5; *SD* = 7.65). A Pearson product-moment correlation was used to assess the relationship between classification accuracy and the number of rules generated by the model. Preliminary analyses showed the relationship to be linear with both variables normally distributed, as assessed by Shaprio-Wilk test ($p > .05$), and there were no outliers. There was no association between classification accuracy and the number of rules present, $r(21) = .056$; $p > .05$. This can clearly be seen in Figure 2. As such, it is beneficial to maintenance operators and decision makers that a smaller set of rules are analysed and understood. This enables a holistic understanding of the underlying behaviour and development of wind turbine pitch faults whilst reducing cognitive load whilst providing comparable classification accuracy to the models with a larger rule base.

Within the model selected, 14 rules were generated leading to an overall classification accuracy of 85.50%. For completeness, the knowledge base determined by the RIPPER algorithm has been included in the appendix. It can be noted that although a high classification accuracy has been attained in this model, it is still difficult to differentiate between no pitch fault existing and a pitch fault being present, with expert analysis required to certify classifications.

As can be seen in Table 3, the Matthews Correlation Coefficient (MCC) (Matthews, 1975) for all classes is strong, showing high correlation between the learnt rules and the validation data. A substantial level of agreement was found between the developed model and the validation data (Cohen's $k = 0.78$; $p < .05$). After deriving the classification, the 14 rules were presented to independent domain experts so that qualitative and quantitative analysis could be performed. Due to the min-max normalization process during pre-processing, values had to be converted back to ensure they were human readable. Once this had been done, a full analysis was performed.

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC	PRC
No Pitch Fault	0.81	0.12	0.77	0.81	0.79	0.68	0.91	0.74
Pitch Fault Developing	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
Pitch Fault Established	0.75	0.01	0.80	0.75	0.78	0.67	0.91	0.79
Weighted Average	0.85	0.07	0.86	0.86	0.78	0.78	0.85	0.85

Table 3 – Descriptive statistics of the developed model.

6. EVALUATION AND EXPLOITATION OF GENERATED RULES

Due to the size of the knowledge base, it was practical to have the domain expert evaluate each rule individually. This is done as the expert can provide a context sensitive ground truth to the analysis, along with experience of situations and conditions which may not have been present within the training data. As domain experts have subjective opinions with regards to what constitutes interesting, novel and important, it is difficult to quantify these characteristics.

However, various artefacts are present within the rule-base which is expected given the nature of the classification. To assess the quality of the rules, a 56- item questionnaire was presented to an independent domain expert who has over 6 years wind turbine diagnostic and prognostic experience within academia. This questionnaire contained a 5-point Likert response scale ranging from 1 (Not intuitive, useful, clear or interesting) to 5 (Highly intuitive, useful, clear or interesting). There were 4 questions presented per rule generated from the model, assessing whether or not the rule was intuitive, useful, clear and interesting

The results of this analysis can be seen in Table 4. As can be seen in Table 4, an average response of 2.89 was recorded; indicating that the rules are typically not particularly intuitive, clear, useful or interesting. This was unexpected. Rules were often regarded as just as useful (*M* = 2.79) as intuitive (*M* = 2.71). This is likely due to the nature of the complex nature of the underlying pitch faults. By having the independent domain expert drive the discussion it was found that of the 14 rules, 11 of the rules were deemed “interesting” and warranted further analysis.

After performing this analysis, the independent domain expert was then presented with a further 13 rules, taken from the work of Kusiak & Verma (2011). To remove potential bias, the expert was not informed of the origin of either set of rules. A 52-item questionnaire was used containing a 5-point Likert scale from 1 (Not intuitive, useful, clear or interesting) to 5 (Highly intuitive, useful, clear or interesting). This was to provide an objective analysis of the intuitiveness, usefulness, clearness and interestingness.

Initially, the expert could not understand the rules due to their format and abstract nature, however, after some time, analysis could be performed. The comparative analysis showed that whilst the rules were found to not be less intuitive (*M* = 1.53; *SD* = 0.63) and clear (*M* = 1.46; *SD* = 0.49), they were still regarded as somewhat useful (*M* = 2.23; *SD* = 0.79) and interesting (*M* = 2.07; *SD* = 0.61). When questioned regarding this, the expert responded that as long as the rules were accurate and accountable, they could be disseminated at a later date. As such, it was determined that an expert system should be developed to aid maintenance operators with enquiries and to handle the large quantities of data present within the system.

6.1. Rule sensitivity to wind turbine location

As different geographical locations have inherently distinct operating conditions, it is expected that the accuracy of the expert system would be reduced when applying the rules to similar wind turbines in a different location. As such, a new expert system would have to be developed for each wind farm as described by the methodology described in Section 4. In order to assess the impact of the geographical location on the accuracy of the rules, data was collected from an additional turbine in the same manner as the previous turbines and was located at a different wind farm within the same country. The wind turbine also had 28 months of SCADA data available, with 3 pitch faults recorded in the historical maintenance log. This wind farm was subject to different external conditions due to being located in a different region.

Validation of the selected model on this additional wind turbine yielded a classification accuracy of 68.68%; somewhat lower than the 85.50% accuracy of the original model. The model was able to identify 2 of the 3 pitch faults which had been recorded in the maintenance log; giving a diagnostic accuracy of 66.67%.

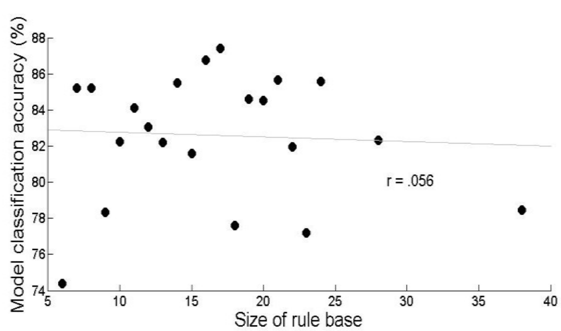

Figure 2. Dominant model classification accuracy plotted against the number of rules generated in each model. No strong correlation existed ($r(21) = 0.06, p > .05$).

7. EXPERT SYSTEM DEVELOPMENT

Due to the strong classification gained from the model, an expert system was developed to aid maintenance managers and decision makers so that available resources could be optimized. Due to the often inaccessible nature of offshore wind turbines, predicting failures can significantly reduce operations and maintenance (OM) costs, thereby increasing the competitive nature of wind energy. The model developed in Section 4 was combined with domain knowledge (meta-data) elicited from the independent domain expert to reduce the high dimensionality of the SCADA data and provide filtering so that the maintenance operator did not have to analyse 190 channels of data coming from over 40 wind turbines per farm, every 10 minutes. In order to assist the operator or decision maker in their role, the expert system must aid their ability to perform analysis and make decisions based upon relevant information. These decisions become more difficult due to various stressors which exist in the working environment. Kontogiannis & Kossiavelou (1999) identify these stressors as:

- Environmental stressors:
  - Noise
  - Temperature
  - Vibrations
- Task Complexity Stressors:
  - Time Pressure
  - Workload
  - Uncertainty
  - Threat/High error consequences
  - Negative feedback
- Group and organisational stressors:
  - Occupational stress
  - Shift/continuous work
  - Lack of team cohesion
  - Communication problems

Due to the nature of the domain, the expert system aims to reduce task complexity stressors. Specifically, reducing time pressure by providing automated analysis and reducing workload by reducing the initial quantity of information presented to the operator per wind turbine. The expert stated that typically, SCADA-alarms for pitch fault are noisy, and only when constant irregularities are noticed over an extended period, is maintenance considered on the turbine.

This is typically due to imperfections within the SCADA system itself causing duplicate, missing and implausible values to be recorded (Sainz, et al 2009). Also, as SCADA data quickly accumulates to create large and unmanageable volumes of data, attempts to deduce the current state of a wind turbine can be severely hindered (Zaher, et al 2009), it is therefore essential that this data can be adequately filtered in an automated manner.

Question	*N*	*M*	*SD*
Intuitive	*14*	*2.71*	*1.09*
Useful	*14*	*2.79*	*0.93*
Clear	*14*	*3.00*	*1.00*
Interesting	*14*	*3.07*	*0.96*

Table 4 – Independent domain expert evaluation.

As such, based upon the expert-knowledge, a threshold was set that should either the “Pitch fault developing” or “Pitch fault established” classification be active for over 90 minutes, an alert would be sent to the maintenance operator. 90 minutes was deemed by the expert to be the minimum length of time an alarm was active before action would be taken and was used as a filter to reduce the noise of the SCADA system. Lower values would increase the noise within the expert system whereas higher values may miss the potential development of pitch faults. This, therefore, reduces the quantity of SCADA alarms presented to the maintenance operator, whilst still presenting those which warranted further investigation.

This reduces the potential cognitive overload of the maintenance operator, allowing for their analysis to be focused on the wind turbines which are current exhibiting potential pitch fault state. This optimises the available maintenance resources by reducing the time spent analysis large quantities of false-positive alarms provided from SCADA system. With regards to the imperfections within the SCADA system, a threshold was also set for missing and implausible values. As missing data cannot fully encapsulate the current operating condition of the wind turbine, it would be difficult to establish if the fault was caused by either a mechanical fault on the turbine, an electrical fault on the turbine or an electrical fault on the SCADA system. As such, 90 minutes of continuous operation in this state provides an alert to the maintenance operator, as above. A similar strategy is employed for implausible data, with expert defined maximum and minimum values for each attribute. Should a single attribute fall outside of this pre-defined range for a full 90 minutes, the operator is also alerted to this.

It should also be noted that one of the alarms on a separate turbine (outside of the training and test data) was active for over 100 days continually. Clearly this is undesirable and hinders the efforts of maintenance managers and decision makers to correctly diagnose and both plan and schedule maintenance.

As such, the ability to correctly filter and classify SCADA-data so that the false-positive instances such as this do not occur is essential. These false-positive instances in the best case are simply a minor hindrance and require further manual analysis by the maintenance operator to determine if a turbine warrants inspection. In the worst case, they provide a basis for maintenance actions which may not be required. In an offshore situation, these un-necessary maintenance actions can be expensive due to the equipment

Turbine	Pitch Fault Alarm Time	Number of Pitch Alarms	Number of Maintenance Jobs	Expert System Alarms	Expert System Time Active
01	15.46 days	193	25	97	10.06
02	17.68 days	222	25	106	12.72
03	12.04 days	27	26	75	8.45
04	19.64 days	215	9	138	9.84

Table 5 – Comparison of Expert System against SCADA-Alarm system

and skills required, and as such, can potentially account for a large portion of maintenance expenditure.

8. Evaluation of Expert System

In order to assess the validity of the expert system developed, historical SCADA-data from 4 wind turbines was used to determine the number of maintenance alerts issued in comparison to the on-board SCADA-alarm system. The validation turbines were independent of those used within the training model, and were located in the same geographical location as the turbines used for the model development and training.

As can be seen in Table 5, in each of the 4 wind turbines analysed, a reduction in the number of alarms generated was observed compared to the turbines integrated SCADA alarm system. This was between 35.80% - 52.26% (M = 44.69%; SD = 6.62%), effectively reducing the workload of the maintenance operator when analysing data to diagnose potential pitch faults. Similarly, this was the case for active alarm time; the reduction was between 28.06% - 49.90% (M = 35.68%; SD = 8.60%). This, again, reduces the quantity of information the maintenance operator has to manage. It is worth noting that although 85 pitch maintenance actions were undertaken over the 28 month period in which this historical data was analysed, 11 of these maintenance actions were not detected by the expert system. This is mainly due to malfunction of the sensors, mechanical systems, and the data collection systems; Of the 11 instances, 7 occurred when data acquisition failed for an extended period. Due to the design of the expert system, missing data does not fully encapsulate the correct turbine condition, and as such, the accuracy is significantly reduced. It is believed that the remaining 4 cases are partly due to time-based preventive maintenance which may not have had sufficient basis for action based upon the observed SCADA-data

9. Conclusions

In this paper we have presented a robust, accurate expert system for the classification and detection of wind turbine pitch faults, as validated by the 85.50% classification accuracy achieved. Transparent, human readable rules were extracted, analysed and verified by an independent domain expert enabling trust in the expert system one of the key barriers to wide scale adoption of CBM technology. These rules were found to be more intuitive than other rules within the literature, and provided the basis for an expert system to aid maintenance operators and decision makers. The number of SCADA alarms was reduced by an average of 44.68%, with a mean reduction of active alarm time by 35.68%. The developed expert system reduced the potential cognitive load on maintenance operators and decision makers by significantly reducing the number of alarms presented to them. This frees maintenance resources, enabling a reduction in annual maintenance costs whilst retaining an equal quality of service. Additionally, no further capital expenditure was necessary due to using pre-existing technological capability. A diagnostic accuracy of 87.05% is achieved in the system, although it is believed that this could be further increased should more reliable sensor technology become available. Our methodology provided a robust strategy to classify SCADA data as having no pitch fault, an established pitch fault or a developing pitch fault. This provides a means to both condition based maintenance and proactive maintenance strategies. By performing remote diagnosis through the expert system, the opportunity for remote maintenance arises due to the nature of the electrical system. In some cases, resetting the control system remedies the existing electrical fault, increasing availability whilst reducing unnecessary maintenance inspections and mitigating the associated costs. By understanding the severity of the fault through the expert system classification, maintenance managers can make informed decisions regarding the most appropriate course of action.

Future work will look to utilise statistical techniques to reduce the quantity of data required for accurate classification. Whilst 4 months of data provided an average classification of 77.29%, had no historical pitch fault data been available, the expert system would not have been able to encapsulate the pitch fault behaviour and would not be fit for purpose. Thus, the expert system would not be effective. As such, the use of suspension histories to classify normal operating behaviour through utilising robust statistical methods would be more appropriate in these circumstances. This would remove the need for fault data present within the training data, providing a strategy for the prognosis and diagnosis of new wind turbines.

APPENDIX

For completeness, the 14 rules learnt by the RIPPER algorithm are presented here. This represents the knowledge base of the expert system.

1. If Alarm is Not Active, and Difference Between Blade Angles is ≤ 18.32 degrees Then Pitch Fault Established.
2. If Alarm is Not Active, and Difference Between Blade Angles is ≤ 18.56 degrees, and Wind Speed ≥ 7.11 m/s Then Pitch Fault Established.
3. If Blade 1 Pitch Motor Torque Maximum ≥ 14.81 kN but ≤ 30.13 kN, and Blade 2 Angle ≥ -12.52 degrees, and Wind Speed ≥ 7.69 m/s, and Then Pitch Fault Established.
4. If Blade 1 Pitch Motor Torque Maximum ≥ 15.59 kN but ≤ 24.35 kN, and Blade 1 Angle ≥ 95.52 degrees, and Wind Speed ≥ 6.73 m/s, and Difference Between Pitch Motor Torques ≤ 41.0 kN Then Pitch Fault Established.
5. If Blade 2 Angle ≤ -0.28 degrees, and Blade 1 Angle ≥ 0.52 degrees, and Wind Speed ≥ 6.44m/s, and Average Pitch Motor Torque ≤ 9.67 kN Then Pitch Fault Established.
6. If Blade 2 Angle ≤ -0.28 degrees, and Blade 1 Angle ≥ -19.74 degrees, and Average Pitch Motor Torque ≤ 10.22 kN, and Wind Speed ≥ 7.42 m/s Then Pitch Fault Established.
7. If Blade 2 Angle ≤ -0.35 degrees, and Blade 1 Angle ≥ -17.13degrees, and Wind Speed ≥ 6.11 m/s, and Difference Between Pitch Motor Torques ≤ 1.08 kN, and Average Pitch Motor Torque ≤ 11.58 kN Then Pitch Fault Established.
8. If Blade 2 Angle ≤ -2.85 degrees, and Blade 1 Angle ≥ -17.13 degrees, and Wind Speed ≥ 7.34 m/s, and Average Pitch Motor Torque ≤ 13.44 kN, Then Pitch Fault Established.
9. If Blade 2 Angle ≤ -2.85 degrees, and Blade 1 Angle ≥ -17.14 degrees, and Wind Speed ≥ 6.19 m/s, and Blade 2 Pitch Motor Torque Maximum ≤ 21.91 kN, and Wind Speed ≥ 6.81 m/s Then Pitch Fault Established.
10. If Blade 2 Angle ≤ -2.98 degrees, and Blade 1 Angle ≥ -17.23 degrees, and Difference Between Pitch Motor Torques ≥ 2.35 kN, and Average Pitch Motor Torque ≤ 10.53 kN, and Wind Speed ≥ 6.56m/s, and Blade 2 Pitch Motor Torque Maximum ≤ 25.02 kN, and Difference Between Blade Angles is ≥ 18.7 degrees Then Pitch Fault Established.
11. If Blade 2 Angle ≤ -3.02 degrees, and Blade 1 Angle ≥ -23.33 degrees, and Wind Speed ≥ 8.25 m/s Then Pitch Fault Established.
12. If Blade 1 Pitch Motor Torque Maximum ≥ 22.58 kN, and Blade 1 Angle ≥ -17.24 degrees, and Wind Speed ≥ 5.80 m/s, and Average Pitch Motor Torque ≤ 10.22 kN, and Blade 2 Angle ≥ -19.08 degrees Then Pitch Fault Established.
13. If Blade 2 Angle ≥ 4.50 degrees Then Pitch Fault Is Developing.
14. Otherwise, No Pitch Fault Is Present.

ACKNOWLEDGMENT

This research is funded through an EPSRC Industrial CASE award in collaboration with 5G Technology Ltd.

REFERENCES

Alpaydin, E. (2004). *Introduction to machine learning*. Cambridge, MA: The MIT Press.

Bianchi, F., De Battista, H., & Mantz, R.J. (2006*). WT Control Systems – Principles, Modelling and Gain Scheduling Design*. London: Springer.

Chen B., Qiu Y., Feng Y., Tavner P., & Song W. (2011). Wind turbine scada alarm pattern recognition. *Renewable Power Generation* (1-6). 6-8 September, Edinburgh, UK.

Cohen, W. W. (1995). Fast effective rule induction. *International conference on machine learning* (115-123). 9-12 July, California, USA.

Cohen W. & Singer Y., (1999). A simple, fast, and effective rule learner," *National. Conference on Artificial Intelligence*. (335 – 342). 18-22 July, Florida, USA.

Crabtree C.J. (2010). *Survey of commercially available condition monitoring systems for wind turbines.* Durham University, 2010.

Djurdjanovic D., Lee J., & Ni J., (2003). Watchdog agent an infotronics-based prognostics approach for product performance degradation assessment and prediction. *Advanced Engineering Informatics*. 17(3), pp. 109–125.

Eti M., Ogaji S., & Probert S. (2006). Reducing the cost of preventive maintenance (pm) through adopting a proactive reliability-focused culture. *Applied Energy* 83(11) pp. 1235-1248.

Garcia V., Sanchez J., Martin-Felez R., & Mollineda R., (2012). Surrounding neighborhood-based smote for learning from imbalanced data sets. *Progress in Artificial Intelligence*, 1(4) pp. 1–16.

Gomez Fernandez J. & Crespo Marquez A. (2009). Framework for implementation of maintenance management in distribution network service providers. *Reliability Engineering & System Safety* 94(10) pp.1639-1649.

Hameed Z., Hong Y., Cho Y., Ahn S., & Song C. (2009). Condition monitoring and fault detection of wind turbines and related algorithms: A review. *Renewable and Sustainable energy reviews*, 13(1) pp. 1–39.

Hatch C. (2004). Improved wind turbine condition monitoring using acceleration enveloping. *GE Energy Journal of Electrical Systems*, 3(1) pp. 26-38.

Jardine A., Lin D., & Banjevic D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7) pp. 1483–1510.

Kim K., Parthasarathy G., Uluyol O., Foslien W., Sheng S., & Fleming P. (2011). Use of SCADA Data for Failure Detection in Wind Turbines. *Energy sustainability and fuel cell science.* ESFuelCell2011-54243. 7-10 August, Washington DC, USA.

Kontogiannis, T., & Kossiavelou, Z. (1999). Stress and team performance: principles and challenges for intelligent decision aids, *Safety Science*. 33(3) pp. 103-128.

Kusiak A & Li W. (2011).The prediction and diagnosis of wind turbine faults.*Renewable Energy*.36(1)pp.16-23.

Kusiak A. & Verma A. (2011). A data-driven approach for monitoring blade pitch faults in wind turbines. *Sustainable Energy, IEEE Trans. on*, 2(1) pp. 87–96.

Levrat E., Iung B., & Marquez A. (2008). E-maintenance: review and conceptual framework. *Production Planning and Control*, 19(4) pp. 408–429.

Lin J. & Zuo M. (2003). Gearbox fault diagnosis using adaptive wavelet filter. *Mechanical systems and signal processing*, 17(6) pp. 1259–1269.

Marais K. & Saleh J. (2009). Beyond its cost, the "value" of maintenance: An analytical framework for capturing its net present value. *Reliability Engineering & System Safety*, 94(2) pp. 644 - 657.

Massoud Amin S. & Wollenberg B. (2005). Toward a smart grid: power delivery for the 21st century. *Power and Energy Magazine, IEEE*, 3(5), pp. 34–41.

Matthews B. (1975). Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochemica et biophysica acta*, 405(2) pp. 442-451.

Moore W. & Starr A. (2006). An intelligent maintenance system for continuous cost-based prioritisation of maintenance activities. *Computers in Industry*, 57(6) pp. 595–606.

Niu G., Yang B., & Pecht M., (2010). Development of an optimized condition-based maintenance system by data fusion and reliability centered maintenance. *Reliability Engineering & System Safety*, 95(7) pp. 786–796.

Quinlan, J. R. (1990). Learning logical definitions from relations. Machine learning, 5(3), pp. 239-266.

Rafiee J., Rafiee M., & Tse P. (2010). Application of mother wavelet functions for automatic gear and bearing fault diagnosis. *Expert Systems with Applications*, 37(6) pp. 4568–4579.

Sainz E,, Llombart A., & Guerrero J. (2009). "Robust filtering for the characterization of wind turbines: Improving its operation and maintenance. *Energy Conversion and Management*, 50(9) pp. 2136–2147.

Wang X. & Makis V. (2009). Autoregressive model-based gear shaft fault diagnosis using the kolmogorov–smirnov test. *Journal of Sound and Vibration*, 327(3), pp.413–423.

Wu S. & Clements-Croome D. (2005). Preventive maintenance models with random maintenance quality. *Reliability Engineering & System Safety*. 90(1) pp. 99-105.

WWEA (2012). "Quarterly bulletin," *World Wind Energy Association Bulletin*, 3(1), October, pp. 1 – 40.

Zaher, S. McArthur, D. Infield, & Y. Patel. (2009). Online wind turbine fault detection through automated scada data analysis. *Wind Energy*, 12(6) pp.574–593.

BIOGRAPHIES

Jamie L. Godwin received the BSc(Hons) degree in computer science from the University Of Durham, England in 2010. At present, he is working towards the degree of Doctor of Philosophy within the department of Engineering and Computing sciences, also at the University of Durham. He has researched areas such as SCADA data analysis and robust multivariate prognostic techniques. He has provided consultancy to various public and private entities, including 5G technologies, Northumbrian Water, various electricity distributors and North Yorkshire Police. His current doctoral research focuses on metrics for maintenance effectiveness, SCADA data analysis and robust multivariate statistical measures for prognosis.

Peter Matthews is a Lecturer in Design Informatics at the School of Engineering and Computing Sciences and holds degrees in Mathematics (1994), Computer Science (1995) and Engineering Design (2002), which were all taken at Cambridge University. He is the author or co-author of numerous books, technical papers and EU patents. His current research interests are centred around industrial data analysis involving collecting and analysing data obtained either from production process monitoring (e.g., SCADA logs from various production machinery) or service life data (e.g. maintenance logs) and utilising Monte Carlo simulations, Evolutionary Algorithms, Bayesian Belief Networks, and interval probabilities (p-boxes) to form future design and operation decisions.